International Journal for Housing Science and Its Applications
Publish August 4, 2025. Volume 46, Issue 3 Pages 1835-1844

https://doi.org/10.70517/ijhsa463145

Extraction of Visual Elements from Dunhuang Frescoes Based on Image Segmentation Algorithm and Application to Modern Visual Communication Design in New Media Environment

Nan Shi[1,2,*] and Yun Qiu[3]
[1] College of Creative Art, Changji University, Changji, Xinjiang, 831100, China
[2] Universiti Teknologi MARA, Creative Art, Shah Alam, Selangor, 40450, Malaysia
[3] Chodang University, 380 Muan-myeon, Muan-eup, Muan-gun, Jeollanam-do, 58530, Republic of Korea
Corresponding authors: (e-mail: yunqiu180@gmail.com).

Abstract Dunhuang murals are artistic cultural heritage with Chinese characteristics, which have rich reference value in visual communication design. In this paper, a fresco segmentation algorithm with enhanced edges is proposed by combining Sobel-Canny edge enhancement and improved GrabCut algorithm. A visual semantic segmentation model based on region suggestion network and full convolutional segmentation network is constructed to realize the high-precision extraction of visual elements of frescoes. The advantages of the proposed model in Dunhuang fresco visual element extraction are analyzed through comparative experiments, and the applicability of the model is explored from the perspective of user experience. The experiments show that the improved GrabCut model outperforms the mainstream segmentation algorithms in terms of PSNR (22.35dB) and SSIM (0.735), and the average running time of the visual semantic segmentation network is only 5.25 seconds. The user survey shows that the proposed model is highly recognized in the three dimensions of functionality, usability, and culture, and the study provides a feasible solution for the digital preservation of cultural heritage and cross-media innovative design.

Index Terms Dunhuang murals, GrabCut algorithm, region suggestion network, full convolutional segmentation network, visual element extraction

I. Introduction

In the desert of northwestern China, Jiuquan City, Gansu Province, lies the world-famous artistic grottoes, the Dunhuang Grottoes [1], [2]. Dunhuang Grottoes consists of 522 caves including Dunhuang Mogao Grottoes, Western Thousand Buddha Caves, and Anxi Yulin Caves, which are the most extant grottoes with the largest number of mural paintings in the world [3], [4]. Dunhuang mural paintings are the treasures of Chinese culture, and the Mogao Caves, which contain Dunhuang mural paintings, are also listed in the World Heritage List by UNESCO [5]. The mural paintings in Mogao Caves are grand in scale and cover a wide range of subjects, from story paintings, landscape paintings, sutra paintings, and traveling pictures, depicting the life scenes and religious culture of ancient people [6]-[8]. As a treasure of Chinese culture, it attracts people all over the world with its unique artistic style and rich cultural connotation [9], [10]. After thousands of years of baptism, Dunhuang murals are still standing, providing people with a window to peep into ancient civilization [11], [12]. How to extract the elements of Dunhuang murals and apply them to modern visual communication design has become a topic of great concern [13].

The extraction and re-creation of the elements of Dunhuang frescoes is a task that is both challenging and meaningful [14]. Under the new media environment, by extracting the lines, colors and characters in Dunhuang murals through image segmentation and other techniques, artists are able to create works with unique styles and far-reaching connotations [15]-[17]. This way of creation not only enables people to better understand and inherit ancient civilization, but also promotes the development of modern visual communication design [18], [19]. The application of Dunhuang murals in modern visual communication design not only demonstrates its profound historical heritage and artistic charm, but also reflects the inheritance and innovation of modern design on traditional culture [20]-[22]. Whether it is brand image design, packaging design or cultural and creative product design, the integration of Dunhuang elements has brought new aesthetic experience and cultural identity to modern life [23], [24].

This paper firstly elaborates the artistry of modeling in Dunhuang frescoes and analyzes the application value of its visual elements. Fusing Sobel and Canny operators to extract the edges of fresco images and combining adaptive wavelet shrinkage algorithm in GrabCut algorithm to realize the segmentation of fresco images. The pre-processed image is segmented by full convolutional segmentation network, and the visual element extraction effect is optimized by relying on region suggestion network. Dice loss function is introduced to improve the model training accuracy. Through a large number of comparison experiments, the feasibility of the improved GrabCut model in the field of mural painting segmentation is verified. Explore the network association advantage of visual semantic segmentation network model from the perspective of operation efficiency and abnormal element detection. Based on the verification of user behavior analysis, examine the application effect of Dunhuang fresco visual elements in modern visual communication design.

II. Extraction of visual elements from Dunhuang murals based on image segmentation algorithm

Dunhuang Mogao Grottoes, as an artistic treasure on the Silk Road, has a high historical, artistic and cultural value, as its murals carry the imprint of Buddhist art, philosophical thinking and national spirit over the past thousand years. However, due to natural weathering, human damage and shooting noise interference, mural images generally have blurred details, color distortion and other problems, restricting its digital protection and innovative applications. With the rapid development of new media technology, how to accurately extract the core elements of mural paintings, such as modeling patterns and color semantics, through computer visual means, and transform them into creative resources for modern visual communication design has become a research hotspot in the field of cultural heritage digitization.

II. A. The artistry of modeling in Dunhuang murals

Dunhuang mural painting through the Silk Road history has a long history, retained a large number of works of art, the works of various dynasties are very artistic value of historical relics. Complex and beautiful shape of the dome algae well pattern is not only rich in variety and color, there are a variety of performance styles and fine painting techniques show the magnificence of traditional Chinese art, it not only has a strong artistic value, but also has a high philosophical value, through the Buddhist stories and paintings and sculpture works to express the spirit of China's nation, for the work of art has given rich vitality.

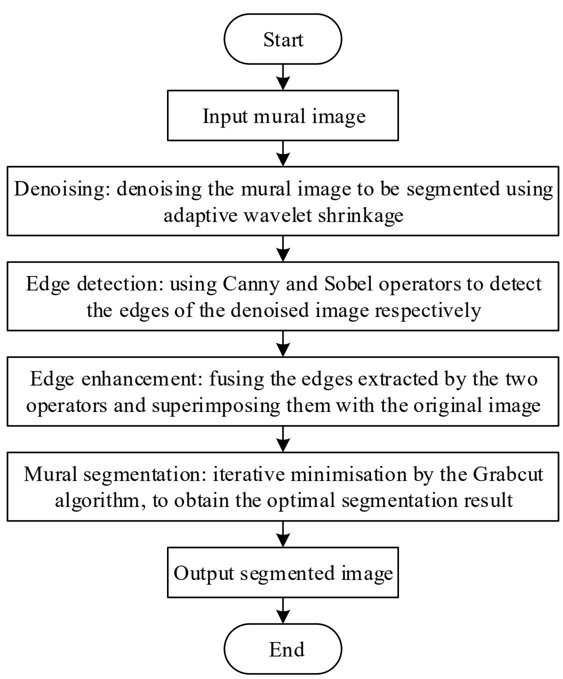

Figure 1: Algorithm process of mural segmentation

In the existing Dunhuang murals may have been a lot can not see the original initial appearance, but from the well-preserved murals can also be seen in its ingenious artistic value, modeling both realistic part of the character's image of the action of the depiction of the vivid image, but also a wealth of subjective part of the creation of artisans according to the story of the scriptures plus their own imaginations, depicted a mysterious color paintings in the development of Chinese culture has had a very important impact on the history of Chinese culture It has had a very important impact on the development of Chinese culture.

II. B.Mural segmentation based on improved GrabCut algorithm

II. B. 1) Mural segmentation process

Adaptive wavelet shrinkage is an improvement of the traditional single-threshold wavelet shrinkage denoising method, which is able to replace the single threshold applied globally with different thresholds suitable for wavelet coefficient amplitudes at all levels of scales, so that the edge information of the image can be better preserved, and there is a better smoothing effect on the noise. Therefore, in this paper, adaptive wavelet shrinkage algorithm is integrated in GrabCut algorithm to eliminate noise in mural images. Based on the respective advantages and disadvantages of Sobel operator and Canny operator, in order to make up for the shortcomings of a single edge detection operator, this paper combines the two operators to extract the edges of the mural image in order to improve the mural segmentation effect. The fusion of wavelet denoising and edge enhancement for Dunhuang fresco segmentation process proposed in this paper is shown in Figure 1.

II. B. 2) Improvements to the GrabCut algorithm

The GrabCut algorithm first requires the user to draw a rectangular box to determine the target region for segmentation, so the initialization of the performance of the GrabCut algorithm is directly affected. The segmentation principle of GrabCut algorithm shows that GrabCut is not good at dealing with noisy images and segmenting the edges, so this paper integrates adaptive wavelet denoising and edge enhancement to improve the GrabCut algorithm.

(1) Adaptive threshold wavelet shrinkage denoising

For the shooting of the mural images of different degrees of noise, this chapter first uses the adaptive threshold wavelet shrinkage denoising method to denoise the mural images, the specific process is as follows:

1) Select the appropriate wavelet basis and wavelet decomposition layers to perform wavelet transform on the image, extract the rapidly changing information in the fresco image as much as possible and minimize the image distortion caused by the number of decomposition layers during wavelet reconstruction.

2) The wavelet decomposition coefficients after wavelet transform are processed with adaptive thresholding to obtain new wavelet coefficients. The decomposed wavelet coefficients have different wavelet coefficient amplitudes at each level of scale and each level of scale is independent of each other, so different matching thresholds are determined according to the different scales at each level, and then the wavelet coefficients at each level of scale are quantized according to the thresholds derived from the corresponding threshold function, so as to obtain the new wavelet coefficients.

3) The new wavelet coefficients are wavelet reconstructed by wavelet inverse transform to obtain the final denoised image.

(2) Sobel and Canny operator edge enhancement

The Sobel and Canny operators adapt to the edges of different images by adjusting the parameters. Threshold is a key parameter in the imaging process of edge detection, and the same operator selects different thresholds for different image scenes. In the algorithm of this paper, the results of edge detection are to enhance the edges of the image and improve the segmentation accuracy of the algorithm, GrabCut algorithm after manually drawing a rectangular box containing the image target, the remaining background information in the box has less impact on the segmentation results, so the threshold is selected to satisfy the target contour of the image is as complete as possible, to reduce the lack of the target edge points, based on which the detection results try to contain fewer of background edge lines is sufficient.

Majority voting algorithm is to find out whether there exists the number of times an element occurs more than $n/2$ in an array containing n elements, which is a simple and effective data fusion method. For the problem of unclear edges of frescoes, combined with the respective characteristics of Sobel operator and Canny operator, this paper adopts the fusion method of majority voting to fuse the edges of fresco images extracted by the two kinds of operators, i.e., after the two kinds of edges are detected and averaged, it is equivalent to forming an array containing 1 element, and if the value of the element is greater than 1/2, then 1 is taken, and vice versa, then 0 to get the complete edge information, the fusion formula is:

$$sum_x = \frac{E_s + E_c}{2} \tag{1}$$

$$I_x = \begin{cases} 1 & if\ sum_x \geq 0.5 \\ 0 & Other \end{cases} \tag{2}$$

where sum_x is the result of the two types of edge detection at any point x in the image summed up and averaged, E_s and E_c represent the values of the edge detection by the Sobel operator and Canny operator at the point x, respectively, and I_x denotes the result of final edge fusion for the point in the image located at x. x point in the image after final edge fusion.

Finally, the fused mural image is superimposed with the original mural image to obtain the edge-enhanced mural image, which is represented as a ternary $T(T_F, T_U, T_B)$ as the input to the GrabCut algorithm by manually drawing a rectangular box. In this case, the user's manual interaction is only required to determine the background region of the image, so it can be assumed that the target region T_F is empty, T_U is the region inside the rectangular box representing the possible target region, and T_B is the region outside the rectangular box representing the background region of the image, i.e., $T_U = overline T_B$.

For the background elements in the image, setting their opacity marker α_n to 0 and the possible target region pixel opacity marker α_n to 1, i.e., $n \in T_B$, there is $\alpha_n = 0$; $n \in T_U$, with $\alpha_n = 1$. The two sets of opacity labels, $\alpha_n = 0$ and $\alpha_n = 1$, are then used to initialize the Gaussian Mixture Model (GMM) of the background and target, respectively.

(3) GrabCut algorithm segmentation

After obtaining the edge-enhanced mural image T, the GrabCut algorithm uses the Gaussian mixing model instead of the statistical histogram model to model the opacity label sets of the target and the background, and updates the label sets as well as the GMM parameters by iterating until the optimal segmentation result is obtained. The process is as follows:

1) According to Eq. (3), find the GMM parameters k_n corresponding to each pixel n in T_U;

$$k_n = \arg\min_{k_n} D_n\left[\alpha_n, k_n, \theta, z_n\right] \tag{3}$$

where D_n is the energy data term corresponding to pixel n, α_n is the opacity scale value corresponding to pixel n, θ is the grayscale histogram of the target or background region of the image, and z_n is the gray scale value corresponding to pixel n.

2) Obtain the GMM parameter θ from the given image data Z according to Equation (4);

$$\theta = \arg\min_{\theta} U\left[\alpha, k, \theta, z\right] \tag{4}$$

where U is the sum of energy data items corresponding to each pixel. α is the opacity scale value, k is the GMM model parameter, θ is the grayscale histogram of the target or background region of the image, and z is an array of grayscale values.

3) According to equation (5), the minimum energy is calculated to get the initial segmentation result;

$$\left\langle \min_{\alpha_n, n \in T_U} \right\rangle \min_k E(\alpha, k, \theta, z) \tag{5}$$

where $E(\alpha, k, \theta, z)$ denotes the Gibbs energy segmented by the GrabCut algorithm.

4) Repeat the execution from step 1) and automatically terminate the iteration to obtain the final segmentation result when there is no significant decay in E.

II. C. Visual element extraction based on visual semantic segmentation network

In order to extract high-precision image visual elements, this paper proposes to combine the region suggestion network with the full convolutional segmentation network to jointly construct a visual semantic segmentation network to complete the extraction of image visual elements.

The structure of visual semantic segmentation network combining region suggestion and full convolutional segmentation network is shown in Figure 2. As can be seen from Fig. 2, firstly, the image is feature extracted by the convolutional layer to generate feature maps, which capture the local and global information of the image. Next, the region suggestion network generates a series of candidate region boxes based on these feature maps, which are designed to quickly localize regions that may contain the target object. To further enhance segmentation accuracy, the feature fusion layer fuses feature maps from different paths, thus integrating multi-scale and multi-level feature information. This step enhances the network's ability to handle complex backgrounds and targets of different sizes. Finally, through an inverse convolution operation, the network restores the fused feature maps to the same size as the original image, achieving pixel-level accurate segmentation. Each pixel is assigned with a corresponding category label, thus completing the visual element extraction task.

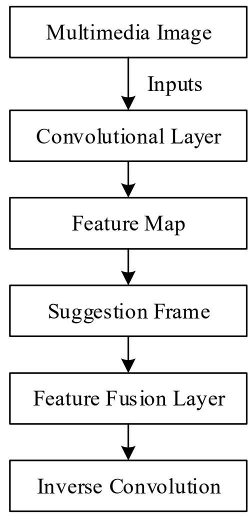

Figure 2: Visual semantic segmentation network structure

For better segmentation of the visual semantic segmentation network during image segmentation, region suggestion frames with region scores are used to assist in its work. This process takes advantage of the fact that the region suggestion frames have less interference from the background image and the higher the region score, the higher the proportion of the image to be segmented will be in the suggestion frame.

A normal distribution method is used to assign weights to the region suggestion frames whose region scores are the first n obtained using the extreme value suppression algorithm. Generally the weights assigned to the center of the image are larger, the weights assigned to the edges of the image are smaller, and the weights are gradually reduced from the value of the center to the surrounding area, as shown in equation (6).

$$G(x,y)=\frac{1}{2}\pi\sigma^2 e^{-\frac{(x^2+y^2)}{2\sigma^2}} \tag{6}$$

In Eq. (6), the positions of the corresponding pixel points in the region suggestion box are expressed using x and y.

In order to obtain the final visual element extraction results through image segmentation, the output image of the full convolutional segmentation network is combined with the weighted region suggestion box, and the output is adjusted to the category probability values of the corresponding pixels in the image through their region scores and weights.

III. Analysis of the Application of Dunhuang Mural Visual Element Extraction in Visual Communication Design

III. A. Improved GrabCut performance level analysis

III. A. 1) Experimental environment and experimental data sources

The experimental environment is based on Window10 operating system, the processor model of the PC is InterCorei7-9750H, the version of the graphics card is NVIDIA GeForce 1660Ti, the experimental platform is JetBrains PyCharm Community Edition 2019, the language is python, and the model is utilized by the TensorFlow deep learning framework, combined with the Keras library to train and test the models in this paper, and the computer vision and machine learning software library Opencv and the annotation software Lableme to process the dataset.

The experimental dataset is divided into training dataset and testing dataset, which contains six different types of labels, including background, animals, houses, auspicious clouds, believers, and Buddha statues, with a total of 500 images from the scanned images of “The Complete Collection of Dunhuang Mural Paintings in China”, and the resize function provided by OpenCV is utilized to change the different types and sizes into the pixels with the resolution of 224×224, and the obtained results are integrated into the original dataset.

The loss function used by the model consists of two parts, the first part is the common cross-entropy loss function, which is used when improving the GrabCut model to classify pixels using the Softmax function; the other part is the Dice loss function, the Dice coefficients are used to compute the similarity of the two samples, and the computational formula is shown in equation (7):

$$s = \frac{2|X \cap Y|}{|X| + |Y|} \tag{7}$$

(The principle of equation (7) is to intersect the predicted results with the true results, multiply by 2 and divide by the sum of the absolute values of the predicted results and the true results. When using the Dice loss function, the positive sample is a small target will produce serious oscillations, when the small target has some pixels of the prediction error, its loss value will have a large change, so the model combined with the cross-entropy loss function. In the training process, every 10epoch is a generation, set the batch_size size to 8, extract batch 300 times in a generation, update the parameters 300 times, use the callback function to supervise the loss value of the training set and the test set, and if the value does not decrease for 3 times, the learning rate is reduced, and if the loss value does not decrease for more than 3 times, then it means that the model training process is over, and the change of the model segmentation accuracy is shown in Fig. 3 shows. Figure 3 shows that the segmentation model in the 1~3 generation, the accuracy of the fast improvement, 3~6 generations of the accuracy of the up and down fluctuation trend, and after 7 generations, the accuracy of the re-convergence of the stability, the model in the 10th generation of the termination of the training, the learning rate to reach the optimal.

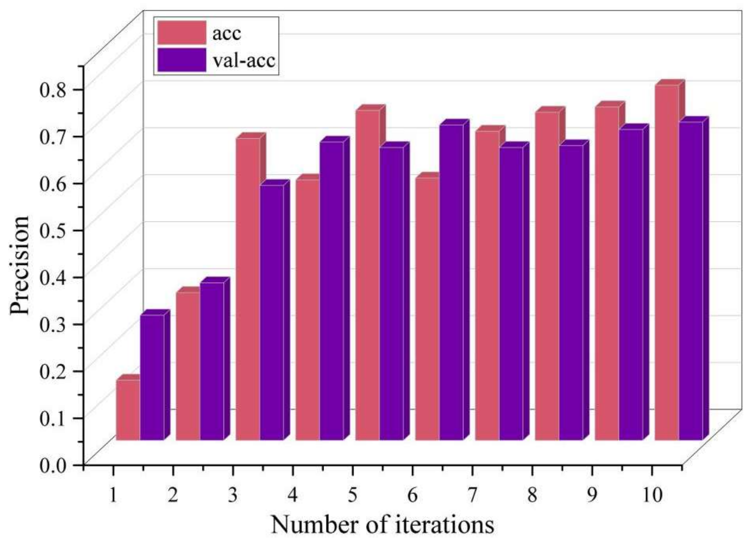

Figure 3: Changes in training accuracy

III. A. 2) Comparative experimental data analysis

In order to verify the excellent characteristics of the improved GrabCut model, the model is evaluated in terms of the model segmentation effect. Firstly, based on the homemade dataset, the SegNet image segmentation model, the PSPNet image segmentation model, the DeeplabV3+ image segmentation model, and the Multi-Classification Lightweight Network Segmentation Model (MC-DM) were selected, and the experimental platform was JetBrains PyCharm Community Edition 2019, and the language was python, utilizing TensorFlow deep learning framework, combined with Keras library to train and test the above models. The traditional segmentation model FCM, GrabCut is selected for testing in MATLABR 2018 platform.

In order to visualize the segmentation effect of each model, 10 different kinds of mural images are randomly selected in the dataset for semantic segmentation, and a single kind of mural image is segmented as the benchmark, and other image elements are used as the background, and the segmentation results are annotated with pixel-level images. Peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) are used as the evaluation indexes of the segmentation results.

PSNR is one of the most common and widely used image evaluation metrics, indicating the ratio of the maximum possible power of the signal to the power of the destructive noise that affects its representation accuracy, with larger values indicating stronger image similarity.Assuming that $x, y \in R^{n\times n}$ two images, where x is a noisy approximation of y, the PSNR is defined as:

$$P(x,y) = 10\lg \frac{255^2}{\frac{1}{mn}\|x-y\|_F^2} \tag{8}$$

The PSNR comparison results of the 10 samples are shown in Table 1. During the experiment, the SegNet model and the PSPNet model perform relatively stable, and the six models have similar segmentation effects when segmenting images with clear outlines and simple constructions like sample 2. While the DeeplabV3+ model shows polarization phenomenon when segmenting images with relatively complex composition like sample 1 and 3, and the segmentation effect is poor for individual images, the MC-DM model and the proposed model perform relatively better, and the average PSNR value of the proposed model reaches 22.35 dB, which is 2.43 dB more than that of the MC-DM model.

Table 1: PSNR(dB)comparison

Sample	1	2	3	4	5	6	7	8	9	10
SegNet	14.35	25.67	16.46	16.03	17.22	18.43	16.35	19.26	20.59	17.55
PSPNet	16.53	23.55	18.94	15.31	17.40	16.53	18.39	16.42	16.94	18.13
DeeplabV3+	7.22	26.18	18.93	16.48	20.19	11.43	18.04	23.53	10.79	15.66
MC-DM	18.22	24.69	20.11	19.32	19.05	19.27	18.33	20.19	20.22	19.78
FCM	8.64	25.74	9.48	16.39	15.24	17.49	11.54	15.42	16.66	12.54
GrabCut	9.03	22.57	8.77	12.53	13.11	12.06	17.03	16.52	20.11	11.75
The proposed	20.93	26.88	21.86	20.75	22.83	21.42	20.97	22.63	23.78	21.45

Since the human eye's visual sensitivity to error is not absolute, its perception results will be affected by the surrounding environment, light sense and many other factors and changes, so there will be a subjective feeling of good but low PSNR value of the phenomenon occurs. To this end, another evaluation index is introduced - structural similarity index SSIM, SSIM is also a kind of index used to measure the similarity of two digital images, compared with PSNR, SSIM in the measurement of the structural quality of the picture is more in line with the human eye for the judgment of the structural quality of the picture, the range of -1 ~ 1, the larger the value, the higher the structural similarity of the picture. The larger the value, the higher the structural similarity of the image. The basic principle of structural similarity is to define the measurement of structural distortion based on the correlation between neighboring pixels, and the sample SSIM comparison results are shown in Table 2. In Table 2, the SSIM value of individual images after DeeplabV3+ model segmentation is slightly decreased, and the SSIM values of other models are consistent with the trend of PSNR values, and the proposed model has the best overall performance, with an average SSIM of 0.735. The combination of the two evaluation indexes shows that the improved GrabCut model segmentation effect is good, with clear edges of the image and good preservation of details, which is applicable to the field of Dunhuang mural segmentation. It is suitable for the segmentation of Dunhuang murals.

Table 2: SSIM comparison

Sample	1	2	3	4	5	6	7	8	9	10
SegNet	0.625	0.846	0.651	0.638	0.663	0.686	0.648	0.688	0.692	0.671
PSPNet	0.644	0.837	0.679	0.622	0.671	0.665	0.673	0.661	0.665	0.672
DeeplabV3+	0.473	0.848	0.683	0.604	0.593	0.618	0.584	0.665	0.593	0.602
MC-DM	0.698	0.852	0.723	0.717	0.709	0.713	0.704	0.735	0.737	0.728
FCM	0.496	0.818	0.502	0.611	0.605	0.621	0.512	0.609	0.622	0.523
GrabCut	0.513	0.823	0.497	0.524	0.531	0.519	0.611	0.602	0.645	0.521
The proposed	0.707	0.868	0.713	0.702	0.735	0.711	0.709	0.733	0.759	0.712

III. B. Analysis of the effect of visual element extraction

In order to verify the efficiency of visual element extraction, the visual element extraction model based on nested increment, visual element extraction model based on SSH framework and visual element extraction model based on visual semantic segmentation algorithm in this paper are used as comparisons for research, respectively. The running time is used as an index to compare the running time of the three different models. The test results are shown in Table 3, in 10 experiments, the average running time of this paper's model is 5.25s, the average time of SSH model is 9.5s, and the average time of nested incremental model is 7.87s. Comparing the average running time of the three different models, it can be seen that the average running time of this paper's model is less than that of the average running time of the SSH model and the nested incremental model.

Table 3: Test results of three different methods

Number of iterations	Running time/s		
	The proposed	SSH	Nested increment
5	4.6	9.2	8.5
10	5.3	9.7	7.6
15	5.8	10.3	8.3
20	4.7	9.5	7.5
25	5.2	9.6	7.7
30	5.9	8.8	7.4
35	4.8	9.1	8.2
40	5.7	8.9	8.1
45	5.2	9.6	7.8
50	5.3	10.3	7.6
Average	5.25	9.5	7.87

When the visual element features are abnormal, the model operation fluctuates, and the abnormal elements are received through the model, and the abnormal elements are examined from the change of the oscillation frequency of the model. The proposed joint model and the visual element extraction model based on region suggestion network and full convolutional segmentation network respectively are used as a comparison for inspection, and the comparison results of the response performance of the three models are shown in Fig. 4. As can be seen from Fig. 4, before receiving abnormal elements, both the proposed model and the comparison model run with insignificant changes in fluctuation amplitude and small changes in oscillation frequency, and both are in a relatively smooth state. After receiving the anomalous elements, the change of oscillation frequency of the two comparison models is very obvious, while the change of oscillation frequency of the proposed model is smaller compared with that before receiving the anomalous elements, and the overall fluctuation amplitude is in a smooth state. The above analysis illustrates that the visual semantic segmentation network model with joint region suggestion and full convolutional segmentation network has better response performance and better visualization than a single model.

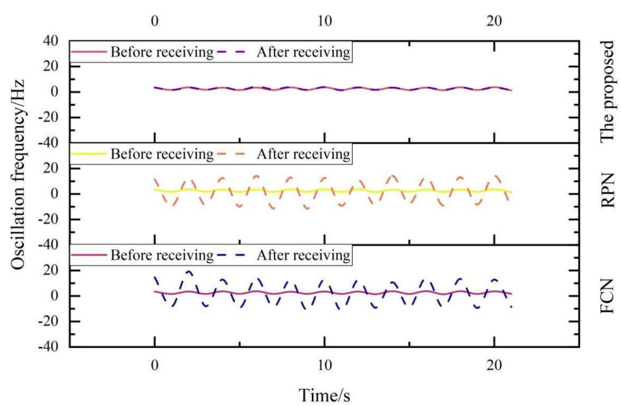

Figure 4: Test results of two different systems

III. C. Application analysis based on user perspective

In order to investigate the application effect of Dunhuang mural visual element extraction in modern visual communication design, this paper selects students majoring in visual communication design in University A to conduct experiments. After a semester of trial, a questionnaire survey was conducted from the perspective of user experience, and a total of 253 questionnaires were received. Before analyzing the data of the questionnaire results, the validity of the questionnaire results should be verified and invalid data should be eliminated to avoid data pollution. After the screening of the questionnaire results, the questionnaire survey received a total of 236 valid questionnaires, the questionnaire validity rate of 93.28%.

The results of user concern elements are shown in Figure 5. Users expressed concerns about extraction efficiency, accuracy, diversity, ease of use, learning, and creativity, all of which accounted for more than 50%. Among them, the extraction efficiency indicator reached more than 90%. This indicates that in the subsequent design process, the design concept of convenience should be upheld to meet user needs.

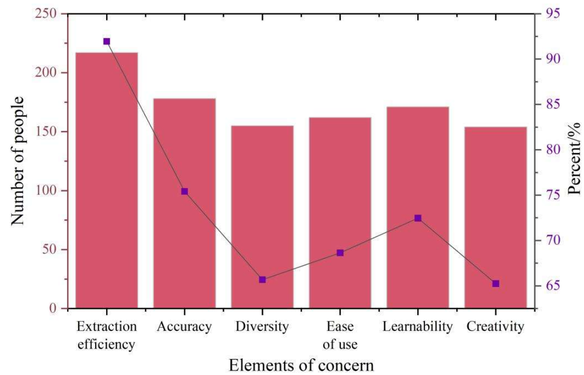

Figure 5: Users care about factor results

The user evaluations were categorized into three categories: functionality, usability, and culture, and subdivided into nine dimensions of elemental diversity, learnability, extraction efficiency, accuracy, ease of use, stability, innovation, creative freedom, and educational value. The nine dimensions correspond to codes d1-d9, and the mean and standard deviation of users in the nine dimensions are shown in Table 4. As can be seen from Table 4, for the design practice of the Dunhuang mural visual element extraction model, the users' evaluations of the nine dimensions in the three categories of functionality, usability, and culture are relatively high. Regarding the mean data, the evaluation scores are basically around 4, with d1, d3, d4, and d5 > 4.5, indicating that users have a high degree of evaluation of the usability experience. The standard deviation reflects the degree of dispersion of a dataset, and the standard deviation values of the evaluation scores are small, between 0.3 and 0.5. This indicates that there is little variability in the ratings among users, reflecting the credibility of the evaluation data results and corroborating the high degree of acceptance of the proposed model among visual communication design students.

Table 4: Statistics of subjective evaluation results under user experience measurement

Evaluation dimension		Average value	Standard deviation
Functionality	d1	4.78	0.32
	d2	4.25	0.37
Usability	d3	4.76	0.41
	d4	4.59	0.33
	d5	4.68	0.35
	d6	4.45	0.42
Cultural nature	d7	4.36	0.33
	d8	4.47	0.45
	d9	4.29	0.38

IV. Conclusion

In this paper, a visual element extraction model based on image segmentation algorithm is designed for the artistry of Dunhuang murals modeling, and the effectiveness of the model and its application in visual communication design is explored through experiments.

The MC-DM model and the improved GrabCut model perform relatively well, and the average PSNR of the improved GrabCut model reaches 22.35 dB, which is 2.43 dB more than that of the MC-DM model.The trend of the SSIM value and the PSNR value is consistent, and the overall performance of the improved GrabCut model is optimal, with an average SSIM of 0.735.In 10 experiments, the average running time of the visual semantic segmentation network model is 0.735, and the average running time of the visual semantic segmentation network model is 0.735. In the 10 experiments, the average running time of the visual semantic segmentation network model is 5.25s, which is less than the average running time of the SSH model and the nested incremental model. The model in this paper combines the region suggestion with the full convolutional segmentation network, and the overall fluctuation amplitude is in a smooth state after receiving anomalous elements, which provides better response performance and better visualization than a single model.

The results of the questionnaire survey show that users express their concerns on extraction efficiency, accuracy, diversity, ease of use, learning, and creativity, with the extraction efficiency indicator reaching more than 90%. Users' evaluations of the nine dimensions in the three categories of functionality, usability, and culture are all relatively high, with evaluation scores of about 4 and standard deviations between 0.3 and 0.5. Among them, d1, d3, d4, and d5 > 4.5 points, indicating that users have a high evaluation of usability experience.

References

[1] Kang, R., & Peters, J. (2019). Dunhuang as a Model for EthnoSTEM Education. The Dunhuang Grottoes and Global Education: Philosophical, Spiritual, Scientific, and Aesthetics Insights, 135-160.
[2] Wang, X., Wang, Y., Guo, Q., Pei, Q., & Zhao, G. (2021). The history of rescuing reinforcement and the preliminary study of preventive protection system for the cliff of Mogao Grottoes in Dunhuang, China. Heritage Science, 9(1), 58.
[3] Li, F., Li, H., & Wang, S. (2024). Quantity and spatial distribution of earth-air activity in Cave 108 of the Dunhuang Mogao Grottoes. Heliyon, 10(4).
[4] Runze, Y. (2023). A study on the spatial distribution and historical evolution of grotto heritage: a case study of Gansu Province, China. Heritage Science, 11(1), 165.
[5] Pi, X. (2022). The connection between Dunhuang Grottoes Painting and Thangka in the middle and late Tang dynasty. Frontiers in Art Research, 4(15).
[6] Hu, G. (2024). The Excavation of the Archaeological Value of Dunhuang Mural Elements has Injected New Vitality into the Research of Modern Animation Image Color Narrative. Mediterranean Archaeology and Archaeometry, 24(1), 333-343.
[7] Wang, S., & Saxdu, I. (2023). The influence of political events and ideology on the formation of the picture concept of Dunhuang Caves Frescos. International Journal of Conservation Science, 14(4), 1443-1462.
[8] Ogura, D., Hase, T., Nakata, Y., Mikayama, A., Hokoi, S., Takabayashi, H., ... & Xue, P. (2021). Influence of environmental factors on deterioration of mural paintings in Mogao cave 285, Dunhuang. Case Studies in Building Rehabilitation, 105-159.
[9] Li, Y., Hu, X., Xie, H., Wu, R., Zhang, Z., Guo, Q., & Hokoi, S. (2024). The impact of cave opening and closure on murals hygrothermal behavior in Cave 98 of Mogao Caves, China. Building and Environment, 256, 111502.
[10] Li, J., Zhang, H., Fan, Z., He, X., He, S., Sun, M., ... & Zhang, B. (2013). Investigation of the renewed diseases on murals at Mogao Grottoes. Heritage Science, 1, 1-9.
[11] Li, H., Wang, W., Zhan, H., Qiu, F., Guo, Q., Sun, S., & Zhang, G. (2017). The effects of atmospheric moisture on the mural paintings of the Mogao Grottoes. Studies in Conservation, 62(4), 229-239.
[12] Yan, Z. (2021, January). The color and artistic features of murals in Dunhuang cave 465 in Mogao Grottoes. In The 6th International Conference on Arts, Design and Contemporary Education (ICADCE 2020) (pp. 56-62). Atlantis Press.
[13] Lu, X., Yang, Y., Wang, L., & Wang, H. (2022). Cultural and creative design of Dunhuang murals in Xixia period based on deep learning. Wireless Communications and Mobile Computing, 2022(1), 7161593.
[14] Cui, Y. (2024). Application of cultural elements of dunhuang murals in landscape design based on mean shift algorithm extraction. Journal of Computational Methods in Science and Engineering, 24(1), 473-487.
[15] Sun, Z., Zhao, D., & Zhang, H. (2022). The innovative application of Dunhuang murals patterns in cultural creation design. Journal of Education, Humanities and Social Sciences, 5, 149-162.
[16] Wang, X., Song, N., Zhang, L., & Jiang, Y. (2018). Understanding subjects contained in Dunhuang mural images for deep semantic annotation. Journal of documentation, 74(2), 333-353.
[17] Liu, F., Yusoff, I. S. M., & Alli, H. (2025). Exploring shape and colors from existing silk scarf by integrating Chinese Dunhuang mural elements to contemporary perception from historical evolution and cultural perspective in China. Herança, 8(1), 164-176.
[18] Sun, M., & Yan, S. (2023, November). Research on the creative design method of Dunhuang mural art based on multi-modal fusion. In Proceedings of the 4th International Conference on Artificial Intelligence and Computer Engineering (pp. 1177-1183).
[19] Qing, L., Zainol, A. S., & Mohd Din, N. H. (2024). Research on Dunhuang cultural and creative product design. Journal of Creative Arts, 1(2), 23-36.
[20] Yin, Y., Sun, D., Su, M., Yu, Z., Su, B., Shui, B., ... & Dong, C. (2019). Investigation of ancient wall paintings in Mogao Grottoes at Dunhuang using laser-induced breakdown spectroscopy. Optics & Laser Technology, 120, 105689.
[21] Sun, T., Jin, T., Huang, Y., Li, M., Wang, Y., Jia, Z., & Fu, X. (2024). Restoring dunhuang murals: crafting cultural heritage preservation knowledge into immersive virtual reality experience design. International Journal of Human–Computer Interaction, 40(8), 2019-2040.
[22] Liu, H., Razak, H. A., Noh, L. M. M., & Lu, X. (2024). From the Ancient Silk Road to Modern Art: Innovative Application of Dunhuang Pattern Art in International Communication. Rita Revista Indexada de Textos Academicos.
[23] Li, K., Puntien, P., & Phuangsua, K. (2024). The applied of China's Dun Huang mural from Fei Tian design in contemporary ceramics product design. FOCUS ON ARTS: FAR, SSRU, 2(2), 34-41.
[24] Zhang, H. (2020). Aesthetic contemplation of the Tang Dynasty Dunhuang frescoes elements on contemporary costume design. In Cross-Cultural Design. User Experience of Products, Services, and Intelligent Environments: 12th International Conference, CCD 2020, Held as Part of the 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings, Part I 22 (pp. 431-440). Springer International Publishing.